

AD-A168 941

ASYMPTOTICALLY OPTIMAL BANDWIDTH SELECTION FOR KERNEL
DENSITY ESTIMATORS F. (U) SOUTH CAROLINA UNIV COLUMBIA
DEPT OF STATISTICS J S MARRON ET AL. MAR 86 TR-113

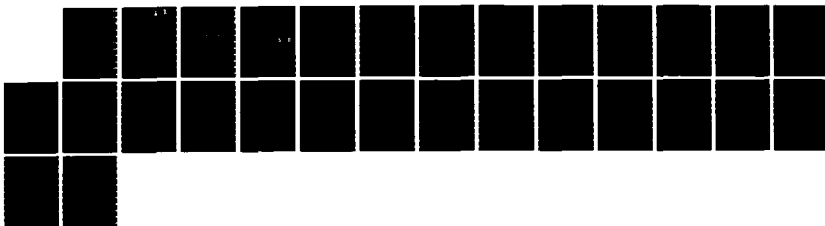
1/1

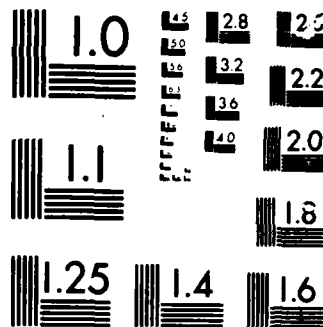
UNCLASSIFIED

AFOSR-TR-86-0277 AFOSR-84-0156

F/G 12/1

NL





5000

DTIC
ELECTE

JUN 10 1986

(2)

UNCLAS

SECURITY CLASS

AD-A168 941

IDENTIFICATION

1a. REPORT SEC

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

3. DISTRIBUTION/AVAILABILITY OF REPORT

Approved for public release,
distribution unlimited

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

Statistics Technical Report No. 113

5. MONITORING ORGANIZATION REPORT NUMBER(S)

AFOSR-TR- 86-0277

6a. NAME OF PERFORMING ORGANIZATION

Department of Statistics

6b. OFFICE SYMBOL
(If applicable)

7a. NAME OF MONITORING ORGANIZATION

Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code)

University of South Carolina
Columbia, SC 29208

7b. ADDRESS (City, State and ZIP Code)

Directorate of Mathematical & Information
Sciences, Bolling AFB, DC 203328a. NAME OF FUNDING/SPONSORING
ORGANIZATION

AFOSR, ARO

8b. OFFICE SYMBOL
(If applicable)

NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

AFOSR-84-0156

8c. ADDRESS (City, State and ZIP Code)

Bolling AFB, DC 20332

10. SOURCE OF FUNDING NOS.

PROGRAM
ELEMENT NO.PROJECT
NO.TASK
NO.WORK UNIT
NO.

61102F

2304

A5

11. TITLE (Include Security Classification)

Asymptotically Optimal Bandwidth Selection for Kernel Density Estimators from Randomly
Right-Censored Samples.

12. PERSONAL AUTHOR(S)

J. S. Marron and W. J. Padgett

13a. TYPE OF REPORT

Technical

13b. TIME COVERED

FROM TO

14. DATE OF REPORT (Yr., Mo., Day)

March, 1986

15. PAGE COUNT

24

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD

GROUP

SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Nonparametric density estimation; Optimal bandwidth;
Random censorship; Smoothing parameter; Cross-validation.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This paper makes two important contributions to the theory of bandwidth selection for kernel density estimators under right censorship. First, an asymptotic representation of the integrated squared error into easily understood variance and squared bias components is given. Second, it is shown that if the bandwidth is chosen by the data-based method of least squares cross-validation, then it is asymptotically optimal in a compelling sense. A by-product of the first part is an interesting comparison of the two most popular kernel estimators.

DTIC

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT

UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION

Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL

Maj. Brian W. Woodruff

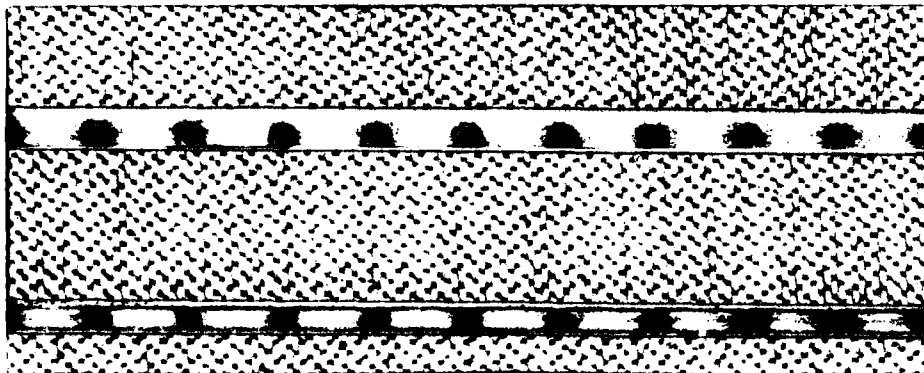
22b. TELEPHONE NUMBER
(Include Area Code)

(202) 765-5027

22c. OFFICE SYMBOL

NM

AFOSR-TR- 86 - 0 2 7 7



DEPARTMENT OF STATISTICS

**The University of South Carolina
Columbia, South Carolina 29208**

*Approved for public release;
distribution unlimited.*

86 6 10 148

ASYMPTOTICALLY OPTIMAL BANDWIDTH SELECTION
FOR KERNEL DENSITY ESTIMATORS FROM RANDOMLY
RIGHT-CENSORED SAMPLES

by

J. S. Marron¹ and W. J. Padgett²

University of South Carolina
Statistics Technical Report No. 113
62G05-15

March, 1986

DTIC
ELECTE
JUN 10 1986
S D D

¹
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

²
Department of Statistics
University of South Carolina
Columbia, SC 29208

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF RESEARCH RESULTS
THIS REPORT IS THE PROPERTY OF THE AFSC
AND IS LOANED TO YOUR ORGANIZATION
DISTRIBUTION IS LIMITED
MATTHEW J. RASCHKE
Chief, Technical Information Division

¹
Research supported by the National Science Foundation Grant DMS-8400602.

²
Research supported by the U. S. Air Force Office of Scientific Research
Grant AFOSR-84-0156 and the U. S. Army Research Office Grant MIPR ARO 139-85.

ABSTRACT

This paper makes two important contributions to the theory of bandwidth selection for kernel density estimators under right censorship. First, an asymptotic representation of the integrated squared error into easily understood variance and squared bias components is given. Second, it is shown that if the bandwidth is chosen by the data-based method of least squares cross-validation, then it is asymptotically optimal in a compelling sense. A by-product of the first part is an interesting comparison of the two most popular kernel estimators.

1. INTRODUCTION

Kernel-type estimators of an unknown probability density function from right-censored data have been studied recently by several authors (e.g. Blum and Susarla, 1980; Diehl and Stute, 1985; Földes, Rejtő and Winter, 1981; McNichols and Padgett, 1985; and Stute, 1985). Padgett and McNichols (1984) gave a review of available results on kernel density estimation from censored data. The details of the forms of these estimators are in section 2.

As in the complete sample (i.e. uncensored) case, the choice of the smoothing parameter, or bandwidth, is crucial to the effective performance of the estimator. Intuitively, if the bandwidth is too small, there is too much "variance" in the sense that features which belong only to the particular data set, and not to the underlying density, may be seen in the estimate. If the bandwidth is too large, there is too much "bias" in the sense that features of the density are smoothed away.

In the complete sample case, an elegant mathematical quantification of the above intuition may be found in Rosenblatt (1956), Parzen (1962), Watson and Leadbetter (1963), and Rosenblatt (1971). In particular, they show that the Mean Integrated Squared Error (MISE) has an asymptotic decomposition as a simple variance term, a simple squared bias term, and some negligible terms. In section 3 it is seen how this type of decomposition may be done in the case of randomly right-censored data. Along the way, approximations are found for the two most popular censored-data kernel estimators which give insight into exactly how they are related.

While this asymptotic representation of MISE provides considerable insight, it is not very useful for selecting the bandwidth because the



Inventory Codes	
Dist	Avail and/or Special
A-1	

minimizer of the two dominant terms contains quantities which are harder to estimate than f itself. As this is also true in the complete sample case, there has recently been considerable work done there on data-based bandwidth selectors. One of the most promising methods is least squares cross-validation, introduced by Rudemo (1982) and Bowman (1984). The bandwidth selected in this way has been shown to be asymptotically optimal under various conditions by Hall (1983), Stone (1984), Burman (1985), Hall (1985), and Marron (1985). Deeper asymptotic properties are established in Hall and Marron (1985a,b).

In section 4, it is shown that least squares cross-validation is also effective in the case of right-censored data. In particular, asymptotic optimality, in the same sense as for the complete sample case, is established. Section 5 contains the proofs. Finally, a practical method for choosing between the two different common kernel estimators is suggested.

2. THE ESTIMATORS

The two best known kernel density estimators are based on estimates of distribution functions. In the censored data case, a widely used distribution function estimator is defined as follows.

Let X_1^0, \dots, X_n^0 denote the i.i.d. survival times of n items or individuals that are censored on the right by i.i.d. random variables U_1, \dots, U_n which are independent of the X_i^0 's. Denote the common distribution function of the X_i^0 's by F^0 and that of the U_i 's by H . Let $H^* = 1-H$. It is assumed that F^0 is absolutely continuous with density f^0 and that H is continuous.

The observed randomly right-censored data are denoted by the pairs

(X_i, Δ_i) , $i=1, \dots, n$, where

$$X_i = \min\{X_i^O, U_i\} \quad \text{and} \quad \Delta_i = 1_{[X_i \leq U_i]},$$

with $1_{[\cdot]}$ denoting the indicator random variable of the event $[\cdot]$.

Based on (X_i, Δ_i) , $i=1, \dots, n$, a popular estimator of the survival function $1-F^O(t)$ is the product-limit (PL) estimator, proposed by Kaplan and Meier (1958) and shown to be "self-consistent" by Efron (1967). Let

(Z_i, Λ_i) , $i=1, \dots, n$, denote the ordered X_i 's along with their corresponding Δ_i 's. The PL estimator of $1-F^O(t)$ is defined by

$$\hat{P}_n(t) = \begin{cases} 1, & 0 \leq t \leq Z_1 \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1} \right)^{\Lambda_i}, & Z_{k-1} < t \leq Z_k, \quad k=2, \dots, n, \\ 0, & t > Z_n. \end{cases}$$

Denote the PL estimator of $F^O(t)$ by $\hat{F}_n(t) = 1 - \hat{P}_n(t)$, and let s_j denote the jump of \hat{P}_n (or \hat{F}_n) at Z_j , that is

$$s_j = \begin{cases} 1 - \hat{P}_n(Z_2), & j=1 \\ \hat{P}_n(Z_j) - \hat{P}_n(Z_{j+1}), & j=2, \dots, n-1 \\ \hat{P}_n(Z_n), & j=n. \end{cases}$$

Then for $j < n$, $s_j = 0$ if and only if $\Lambda_j = 0$, that is, Z_j is a censored observation. For various properties of the PL estimator, see Breslow and Crowley (1974), Csörgő and Horváth (1983), Földes and Rejtő (1981), Földes, Rejtő and Winter (1980), Gill (1983), and Wellner (1982), among others.

The distribution function estimator, \hat{F}_n , is very naturally used to construct a density estimator by defining

$$\begin{aligned} f_n(x) &= h^{-1} \int_{-\infty}^{\infty} K\left(\frac{x-t}{h}\right) dF_n(x) \\ &= h^{-1} \sum_{j=1}^n s_j K\left(\frac{x-Z_j}{h}\right). \end{aligned}$$

This estimator has been studied by Foldes, Rejto and Winter (1981), McNichols and Padgett (1986), Diehl and Stute (1985), and Stute (1985).

An alternative kernel estimator has been proposed by Blum and Susarla (1980), extending the results of Rosenblatt (1976) to censored data. It is motivated by the fact that a reasonable (and technically easy to handle) estimate of $f^0(x)H^*(x)$ is given by

$$(f^{0H^*})_n(x) = (nh)^{-1} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) 1_{[\Delta_j=1]}.$$

Hence, it makes sense to estimate $f^0(x)$ by $(f^{0H^*})_n(x)$ divided by an estimate of $H^*(x)$. If we reverse the intuitive roles played by X_i^0 and U_i , then the product-limit estimator for H^* is given by

$$\hat{H}_n(t) = \begin{cases} 1, & 0 \leq t \leq Z_1 \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1}\right)^{1-\Lambda_i}, & Z_{k-1} < t \leq Z_k, k=2, \dots, n \\ 0, & t > Z_n. \end{cases}$$

This does not make a good denominator because it takes on the value zero, so Blum and Susarla propose changing \hat{H}_n slightly to

$$H_n^*(t) = \begin{cases} 1, & 0 \leq t \leq Z_1, \\ \prod_{i=1}^{k-1} \left(\frac{n-i+1}{n-i+2}\right)^{1-\Lambda_i}, & Z_{k-1} < t \leq Z_k, k=2, \dots, n, \\ \prod_{i=1}^n \left(\frac{n-i+1}{n-i+2}\right)^{1-\Lambda_i}, & Z_n < t. \end{cases}$$

Hence, define

$$f_n^*(x) = [nhH_n^*(x)]^{-1} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) 1_{[\Delta_j=1]} .$$

To get some idea for what the relationship is between the estimators f_n and f_n^* , note that from Susarla, Tsai, and Van Ryzin (1984) for each j , $s_j = \Lambda_j [\hat{nH}_n(Z_j)]^{-1}$. Hence, we may write:

$$f_n(x) = \sum_{j=1}^n \frac{\Delta_j}{\hat{nH}_n(X_j)h} K\left(\frac{x-X_j}{h}\right), \quad (2.1)$$

$$f_n^*(x) = \sum_{j=1}^n \frac{\Delta_j}{nH_n^*(x)h} K\left(\frac{x-X_j}{h}\right). \quad (2.2)$$

Since \hat{H}_n and H_n^* are essentially the same, the only significant difference between the estimators is the argument of the estimate of H^* . It will be seen in the next section that the difference is typically not negligible.

It will be assumed throughout that K is a probability density with compact support and that K is Hölder continuous. In addition, $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$. Letting $T_G = \sup\{t: G(t) < 1\}$ for a distribution function G , it is assumed that $T_H \leq T_{F_0} \leq \infty$ and that f^{OH^*} is Hölder continuous of order $\varepsilon > 0$.

3. ASYMPTOTIC REPRESENTATION

The main idea of this section is that $f_n(x)$ and $f_n^*(x)$ are essentially the same as

$$\tilde{f}_n(x) = \sum_{j=1}^n \frac{\Delta_j}{nH^*(X_j)h} K\left(\frac{x-X_j}{h}\right), \quad (3.1)$$

$$\tilde{f}_n^*(x) = \sum_{j=1}^n \frac{\Delta_j}{nH^*(x)h} K\left(\frac{x-X_j}{h}\right),$$

respectively, because the convergence of \hat{H}_n and H_n^* to H^* is faster ($\sim n^{-\frac{1}{2}}$)

than that of the density estimators (often $\sim n^{-2/5}$). Essentially, the same idea has been used by Diehl and Stute (1985) and Stute (1985). For \hat{f} equal to any of f_n , f_n^* , \bar{f}_n , or \bar{f}_n^* , we choose to analyze its performance by studying the Integrated Squared Error, $ISE(\hat{f}) = \int_0^\infty [\hat{f}(x) - f^0(x)]^2 w(x) dx$, where $w(x)$ is a nonnegative weight function.

There are two major reasons for working with ISE instead of with its expected value, MISE. First, ISE is a more compelling error criterion because it assesses how well \hat{f} is doing for the data set at hand, instead of only for the average over all possible data sets as is done by MISE. Second, ISE is more natural for the automatic bandwidth selection results of the next section. It should be pointed out that by using methods slightly easier than those used here, all of our results can be formulated in terms of MISE. Also, there is an obvious extension of the theorems of this section to the pointwise convergence of the estimators when it is assessed by the Mean Square Error.

The role of the weight function, w , is to eliminate endpoint effects. Assume in particular that w is bounded and supported on $[0, T]$, where $T < T_H$.

The statement of the theorem will be uniform over $h \in [n^{-1+\epsilon}, n^{-\epsilon}]$, some $\epsilon > 0$. This is necessary for the automatic bandwidth selection results of section 4.

Theorem 3.1. Under the conditions on K , $f^0 H^*$, F^0 , f^0 , and H stated in section 2, for $h \in [n^{-1+\epsilon}, n^{-\epsilon}]$, we have

$$\begin{aligned} \sup_h \left| \frac{ISE(f_n(x)) - [an^{-1}h^{-1} + b]}{an^{-1}h^{-1} + b} \right| &\rightarrow 0 \text{ a.s.,} \\ \sup_h \left| \frac{ISE(f_n^*(x)) - [an^{-1}h^{-1} + b^*]}{an^{-1}h^{-1} + b^*} \right| &\rightarrow 0 \text{ a.s.,} \end{aligned} \quad (3.2)$$

where

$$a = (\int K^2) (\int \frac{f^0 w}{H^*}),$$

and where b, b^* are defined by

$$b = \int B(x, h)^2 w(x) dx,$$

$$b^* = \int B^*(x, h)^2 \frac{w(x)}{[H^*(x)]^2} dx,$$

$$B(x, h) = \int K(u) [f^0(x-hu) - f^0(x)] du,$$

$$B^*(x, h) = \int K(u) [f^0(x-hu) H^*(x-hu) - f^0(x) H^*(x)] du.$$

Remark 3.1. Note that an immediate consequence of Theorem 3.1 is the ISE consistency of f_n and f_n^* .

Remark 3.2. The only difference in the asymptotic representations of ISE shows up in the bias part. Note that for some choices of f^0 and H^* , b will be smaller, while for other choices, b^* will be smaller. Hence, the estimators f_n and f_n^* are really not comparable from this representation. However, note that, by an addition-subtraction,

$$\frac{B^*(x, h)}{H^*(x)} = \int K(u) f^0(x-hu) \left[\frac{H^*(x-hu) - H^*(x)}{H^*(x)} \right] dx + B(x, h).$$

So in a weak sense, f_n^* has an extra "noise term", which may make f_n slightly preferable.

Rates of convergence may be computed in the usual manner of Rosenblatt and Parzen. Further, Theorem 3.1 yields an asymptotic bandwidth which is optimal in the same sense as the bandwidths of Rosenblatt and Parzen except that the random error criterion ISE is used in place of its mean. This is given in the next remark.

Remark 3.3. (i) It is well known in the complete sample case that by allowing K to take on negative values, a faster rate of convergence can be obtained.

Theorem 3.1 demonstrates that the same is true here. In particular, suppose

$$\int x^j K(x) dx = \begin{cases} 1, & j=0, \\ 0, & j=1, \dots, k-1, \\ \kappa, & j=k, \end{cases} \quad (3.3)$$

(for $k > 2$, this violates the assumptions of Theorem 3.1; however it is straightforward, but space-consuming, to modify the proofs to allow for this). If we assume that f^0 and f^{0H^*} have k uniformly continuous derivatives, then

$$b = h^{2k} \left(\frac{\kappa}{k!} \right)^2 \int [(f^0)^{(k)}]^2 w dx + o(h^{2k}),$$

$$b^* = h^{2k} \left(\frac{\kappa}{k!} \right)^2 \int [(f^{0H^*})^{(k)}]^2 \frac{w}{(H^*)^2} dx + o(h^{2k}).$$

Hence, for the estimator f_n , the "classical optimal bandwidth" has the form

$$h_0 = \left\{ \frac{(\int K^2) \left(\int \frac{f^0 w}{H^*} \right)}{\left(\frac{\kappa}{k!} \right)^2 \int [(f^0)^{(k)}]^2 w} \right\}^{1/(2k+1)} n^{-1/(2k+1)},$$

and the rate of convergence is $ISE \sim n^{-2k/(2k+1)}$. Here and in the following remarks, there are obvious analogues for the estimator f_n^* .

To see how Theorem 3.1 implies that h_0 behaves like the optimal bandwidth of Rosenblatt and Parzen (the complete sample case), define

$$EI_0 = n^{-1} h^{-1} \left[\int K^2 \right] \left[\int \frac{f^0 w}{H^*} \right] + h^{2k} \left(\frac{\kappa}{k!} \right)^2 \int [(f^0)^{(k)}]^2 w.$$

By (3.2), with obvious notation,

$$\sup_h \left| \frac{ISE(f_n, h) - EI_0(h)}{EI_0(h)} \right| \rightarrow 0 \quad \text{a.s.}$$

Let h_M denote the minimizer of $ISE(f_n, h)$ and recall that h_0 is the minimizer of $EI_0(h)$. Then from the inequalities $ISE(f_n, h_0) \geq ISE(f_n, h_M)$ and $EI_0(h_M) \geq EI_0(h_0)$, it follows that

$$\begin{aligned}
\left| \frac{\text{ISE}(f_n, h_0) - \text{ISE}(f_n, h_M)}{\text{ISE}(f_n, h_0)} \right| &\leq \left| \frac{\text{ISE}(f_n, h_0) - \text{EI}_0(h_0)}{\text{EI}_0(h_0)} \right| \frac{\text{EI}_0(h_0)}{\text{ISE}(f_n, h_0)} \\
&+ \left| \frac{\text{ISE}(f_n, h_M) - \text{EI}_0(h_M)}{\text{EI}_0(h_M)} \right| \frac{\text{EI}_0(h_M)}{\text{ISE}(f_n, h_M)} \\
&\rightarrow 0 \quad \text{a.s.}
\end{aligned}$$

Hence,

$$\frac{\text{ISE}(f_n, h_0)}{\inf_h \text{ISE}(f_n, h)} \rightarrow 1 \quad \text{a.s.},$$

which shows that h_0 is optimal in the same sense as the bandwidths of Rosenblatt and Parzen, except for the fact that the random ISE criterion is used in place of its mean.

Remark 3.3. (ii) If we keep the assumption (3.3), but suppose f^0 has $p < k$ derivatives (p need not be an integer by putting a Hölder condition of order $p - [p]$ on the $[p]$ -th derivative, where $[\cdot]$ denotes the greatest integer less than or equal to p), then it can be shown that $b^* \leq C h^{2p}$ for some positive constant C . Hence, by taking $h \sim n^{-1/(2p+1)}$, the well-known (see, for example, Bretagnolle and Huber (1979)) "optimal rate," $\text{ISE} \sim n^{-2p/(2p+1)}$, can be obtained for our censored data problem.

4. AUTOMATIC BANDWIDTH SELECTION

For data-based bandwidth selection, we propose least-squares cross-validation, which was invented, for complete sample density estimators, by Rudemo (1982) and Bowman (1984). This is motivated as follows. Let \hat{f} denote either f_n or f_n^* . Since the third term of

$$\text{ISE}(\hat{f}) = \int \hat{f}^2 w - 2 \int \hat{f} f^0 w + \int (f^0)^2 w$$

is independent of h , we would like to choose h to minimize the sum of the first two terms. The first term is known. The integral of the second term can

be unbiasedly estimated by

$$n^{-1} \sum_{i=1}^n \hat{f}_i(X_i) \frac{w(X_i)}{H_n^*(X_i)} 1_{[\Delta_i=1]},$$

where \hat{f}_i is the "leave-one-out" version of \hat{f} , given by

$$f_{n,i}(x) = \sum_{j \neq i} \frac{1}{(n-1)H_n^*(X_j)h} K\left(\frac{x-X_j}{h}\right) 1_{[\Delta_j=1]}$$

when \hat{f} is f_n , and by

$$f_{n,i}^*(x) = \sum_{j \neq i} \frac{1}{(n-1)H^*(x)h} K\left(\frac{x-X_j}{h}\right) 1_{[\Delta_j=1]}$$

when \hat{f} is f_n^* . Thus, we define \hat{h}_c to be the minimizer of the least-squares cross-validation criterion

$$CV(h) = \int [\hat{f}(x)]^2 w(x) dx - 2n^{-1} \sum_{i=1}^n \hat{f}_i(X_i) \frac{w(X_i)}{H_n^*(X_i)} 1_{[\Delta_i=1]}.$$

Theorem 4.1. Under the conditions of Theorem 3.1, \hat{h}_c is asymptotically optimal in the sense that

$$\frac{ISE(\hat{f}, \hat{h}_c)}{\inf_h ISE(\hat{f}, h)} \rightarrow 1 \quad \text{a.s.}$$

Remark 4.1. Theorem 4.1 says that \hat{h}_c is optimal under either of the assumptions stated in Remark 3.3 (i) or Remark 3.3 (ii). This generalizes the important asymptotic optimality results of Hall (1983), Stone (1984), Burman (1985), Hall (1985), and Marron (1985) to the case of censored data.

Remark 4.2. The fact that $CV(h)$ essentially provides an estimate of $ISE(\hat{f}, h)$ suggests a practical method of choosing between f_n and f_n^* . In particular, if $CV(h)$ for $\hat{f} = f_n$ is smaller than $CV(h)$ for $\hat{f} = f_n^*$, then the estimator f_n should be used, as its ISE will probably be smaller.

5. PROOFS OF THEOREMS

All proofs are given for the estimator $f_n(x)$, as it will be obvious how to adapt them to handle $f_n^*(x)$. The symbol C will be used for a generic constant. Note first that, using the notation (3.1), by adding and subtracting $\bar{f}_n(x)$,

$$\text{ISE}(f_n) = \text{ISE}(\bar{f}_n) + \text{II} + \text{III}, \quad (5.1)$$

where

$$\text{II} = 2 \int_0^\infty [\bar{f}_n(x) - f^0(x)][f_n(x) - \bar{f}_n(x)] w(x) dx,$$

$$\text{III} = \int_0^\infty [f_n(x) - \bar{f}_n(x)]^2 w(x) dx.$$

Proof of Theorem 3.1. We analyze each of the terms $\text{ISE}(\bar{f}_n)$, II, and III separately. First, by a "variance-bias squared" decomposition, and standard computations of the type in Rosenblatt (1971),

$$\text{MISE}(\bar{f}_n) = E(\text{ISE}(\bar{f}_n)) = v + b, \quad (5.2)$$

where

$$v = n^{-1} h^{-1} (\int K^2) (\int \frac{f^0 w}{H}) + o(n^{-1} h^{-1}), \quad (5.3)$$

and where b is defined in section 3. The fact that $\text{ISE}(\bar{f}_n)$ behaves like $\text{MISE}(\bar{f}_n)$ is contained in the following lemma.

Lemma 1.
$$\sup_h \left| \frac{\text{ISE}(\bar{f}_n) - \text{MISE}(\bar{f}_n)}{\text{MISE}(\bar{f}_n)} \right| \rightarrow 0 \quad \text{a.s.}$$

The fact that term III is negligible is contained in

Lemma 2.
$$\sup_h \left| \frac{\text{III}}{\text{MISE}(\bar{f}_n)} \right| \rightarrow 0.$$

It follows from the Schwartz Inequality, Lemma 1, and Lemma 2 that III may be replaced by II in the statement of Lemma 2.

This last fact, together with (5.1), (5.2), (5.3), Lemma 1, and Lemma 2

complete the proof of Theorem 3.1.

Before proving Theorem 4.1, we give the proof of Lemmas 1 and 2.

Proof of Lemma 1. Let $N = \#(\Delta_i=1)$. For $v=1, \dots, n$, conditioning on $[N=v]$, $\{X_i: \Delta_i=1\}$ is a set of v i.i.d. random variables with density $f^0 H^*/p$, where

$$p = \int_0^\infty f^0(x) H^*(x) dx.$$

Let E_v denote expectation under this conditional distribution. The method of the proof of Theorem 1 of Marron and Härdle (1986) shows that, under the stated assumptions, for $k=1, 2, \dots$, there exist constants $C > 0$ and $\gamma > 0$ so that

$$\sup_h E_v \left[\frac{ISE(\bar{f}_n) - E_v(ISE(\bar{f}_n))}{E_v(ISE(\bar{f}_n))} \right]^{2k} \leq C v^{-\gamma k}. \quad (5.4)$$

To analyze $E_v(ISE(\bar{f}_n))$, note first that

$$\begin{aligned} E_v \bar{f}_n(x) - f^0(x) &= \\ &= \frac{v}{n} \int \frac{1}{H^*(y)h} K\left(\frac{x-y}{h}\right) \frac{f^0(y)H^*(y)}{p} dy - f^0(x) \\ &= \int K(u) \left[\frac{v}{np} f^0(x-hu) - f^0(x) \right] du \\ &= \frac{v}{np} B(x, h) + \left(\frac{v}{np} - 1 \right) f^0(x), \end{aligned}$$

where $B(x, h)$ was defined in section 3. Next note that

$$\begin{aligned} E_v [\bar{f}_n(x) - E_v \bar{f}_n(x)]^2 &= \text{var}_v \left[\sum_{i=1}^n \frac{1}{nH^*(X_i)h} K\left(\frac{x-X_i}{h}\right) 1_{[\Delta_i=1]} \right] \\ &= \frac{v}{n^2} \text{var}_v \left[\frac{1}{H^*(X_i)h} K\left(\frac{x-X_i}{h}\right) \right] \\ &= \frac{v}{np} n^{-1} h^{-1} (\int K^2) \frac{f^0(x)}{H^*(x)} + o\left(\frac{v}{np} n^{-1} h^{-1}\right). \end{aligned}$$

Thus, by a "variance-bias squared" decomposition,

$$E_v(ISE(\bar{f}_n)) = v_v + b_v.$$

where

$$v_v = \frac{v}{np} v + o\left(\frac{v}{np} n^{-1} h^{-1}\right),$$

for v defined in (5.3), and where

$$\begin{aligned} b_v &= \left(\frac{v}{np}\right)^2 b + 2\left(\frac{v}{np}\right)\left(\frac{v}{np} - 1\right) \int_0^\infty B(x, h) f^0(x) w(x) dx \\ &\quad + \left(\frac{v}{np} - 1\right)^2 \int_0^\infty f^0(x)^2 w(x) dx, \end{aligned}$$

for b as in section 3. Hence,

$$\begin{aligned} E_v(ISE(\tilde{f}_n)) &= MISE(\tilde{f}_n) + \left(\frac{v}{np} - 1\right)v + o\left(\frac{v}{np} n^{-1} h^{-1}\right) \\ &\quad + \left(\left(\frac{v}{np}\right)^2 - 1\right)b + 2 \frac{v}{np} \left(\frac{v}{np} - 1\right) \int_0^\infty B(x, h) f^0(x) w(x) dx \\ &\quad + \left(\frac{v}{np} - 1\right)^2 \int_0^\infty f^0(x)^2 w(x) dx. \end{aligned}$$

Now for small $\tau > 0$ and for $n=1, 2, 3, \dots$, restrict attention to v between $np - n^{\frac{1}{2}+\tau}$ and $np + n^{\frac{1}{2}+\tau}$. For such v , $\frac{v}{np} \leq 2$ and

$$\left| \frac{v}{np} - 1 \right| \leq C n^{-\frac{1}{2}+\tau}$$

for a constant C . It follows from (5.2) and (5.3) that, for a different value of C , and for n sufficiently large,

$$\inf_h MISE(\tilde{f}_n) \geq C n^{-1+\epsilon}. \quad (5.5)$$

Hence, for small τ , large n , and another C ,

$$\sup_h \left| \frac{E_v(ISE(\tilde{f}_n)) - MISE(\tilde{f}_n)}{MISE(\tilde{f}_n)} \right| \leq C n^{-\epsilon+2\tau}.$$

Thus, for such v , from (5.4),

$$\sup_h E_v \left[\left| \frac{ISE(\tilde{f}_n) - MISE(\tilde{f}_n)}{MISE(\tilde{f}_n)} \right|^{2k} \right] \leq C n^{-\gamma k}.$$

Now, let Γ_n be a subset of $[n^{-1+\epsilon}, n^{-\epsilon}]$ so that successive members of Γ_n are separated by a distance less than or equal to $n^{-\rho}$ and so that $\#(\Gamma_n) \leq n^\rho$ for some $\rho > 0$. Then, using obvious notation,

$$\begin{aligned}
& P\left(\sup_h \left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} \right| \geq \varepsilon\right) \\
& \leq P\left[\sup_{h \in \Gamma_n} \left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} \right| > \frac{\varepsilon}{2}\right] \\
& + P\left[\sup_{|h-h'| \leq n^{-p}} \left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} - \frac{\text{ISE}(\tilde{f}_n, h') - \text{MISE}(\tilde{f}_n, h')}{\text{MISE}(\tilde{f}_n, h')} \right| > \frac{\varepsilon}{2}\right] \\
& = \sum_{v=0}^n \binom{n}{v} p^v (1-p)^{n-v} P_v \left[\sup_{h \in \Gamma_n} \left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} \right| > \frac{\varepsilon}{2} \right],
\end{aligned}$$

where the last equality comes from a continuity argument and the assumptions that K is Hölder continuous and has compact support. Letting

$$A_{n,\tau} = [np - n^{\frac{1}{2}+\tau}, np + n^{\frac{1}{2}+\tau}],$$

$$\begin{aligned}
& P\left[\sup_h \left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} \right| > \varepsilon\right] \\
& \leq \sum_{v \in A_{n,\tau}} \binom{n}{v} p^v (1-p)^{n-v} P_v \left[\sup_h \left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} \right| > \frac{\varepsilon}{2} \right] \\
& \quad + \sum_{v \notin A_{n,\tau}} \binom{n}{v} p^v (1-p)^{n-v} \\
& \leq \sum_{v \in A_{n,\tau}} \binom{n}{v} p^v (1-p)^{n-v} \sum_{h \in \Gamma_n} P_v \left[\left| \frac{\text{ISE}(\tilde{f}_n, h) - \text{MISE}(\tilde{f}_n, h)}{\text{MISE}(\tilde{f}_n, h)} \right| > \frac{\varepsilon}{2} \right] + 2\Phi(-n^\tau)
\end{aligned}$$

$$\leq \sum_{v \in A_{n,\tau}} \binom{n}{v} p^v (1-p)^{n-v} n^{\rho(\frac{2}{\varepsilon}) 2k} \sup_h E_v \left[\frac{ISE(\tilde{f}_n, h) - MISE(\tilde{f}_n, h)}{MISE(\tilde{f}_n, h)} \right]^{2k} + 2\phi(-n^\tau) \\ \leq C n^\rho n^{-\gamma k} + 2\phi(-n^\tau), \quad (5.6).$$

where ϕ denotes the standard normal cdf. But, for k sufficiently large, the first term on the right side of (5.6) is summable on n , and, since the second term is also summable on n , the proof of Lemma 1 is complete.

Proof of Lemma 2. Using the assumption on the support of w , and using the compactness of the support of K , observe that for n sufficiently large,

$$\sup_h III = \sup_h \int_0^\infty \left[\sum_{i=1}^n \left(\frac{1}{H_n(X_i)} - \frac{1}{H^*(X_i)} \right) \frac{1}{nh} K\left(\frac{x-X_i}{h}\right) 1_{[\Delta_i=1]} \right]^2 w(x) dx \\ \leq \left(\sup_{t \in [0, T']} \left| \frac{1}{\hat{H}_n(t)} - \frac{1}{H^*(t)} \right| \right) \left(\sup_h \int_0^\infty [(f^{\circ H^*})_n(x)]^2 w(x) dx \right),$$

where $T' = (T+T_H)/2$, and where $(f^{\circ H^*})_n$ was defined in section 2. Lemma 2 is now a consequence of the results of Csörgő and Horváth (1983) together with (5.5) and the fact that there is a constant C so that

$$\sup_h \int_0^\infty [(f^{\circ H^*})_n(x)]^2 w(x) dx \leq C \quad \text{a.s.} \quad (5.7)$$

To verify (5.7), note that by adding and subtracting $f^{\circ}(x)H^*(x)$,

$$\int_0^\infty [(f^{\circ H^*})_n(x)]^2 w(x) dx = U + V + W,$$

where

$$U = \int_0^\infty [(f^{\circ H^*})_n - f^{\circ H^*}]^2 w(x) dx,$$

$$V = 2 \int [(f^0 H^*)_n - f^0 H^*] [f^0 H^*] w(x) dx,$$

$$W = \int [f^0 H^*]^2 w(x) dx.$$

Now W is deterministic and independent of h . An argument similar to (but slightly easier than) that used above on $ISE(\tilde{f}_n)$ gives

$$\sup_h U \rightarrow 0 \quad \text{a.s.}$$

An application of the Schwartz inequality to V yields (5.7), which completes the proof of Lemma 2.

Proof of Theorem 4.1. Here again, only the proof in the slightly harder case of $\hat{f} = f_n$ is given. We note that by a computation similar to that used to verify Remark 3.3 (i), Theorem 4.1 follows from (3.2) and the result that

$$\sup_{h, h'} \frac{|CV(h) - ISE(f_n, h) - [CV(h') - ISE(f_n, h')]|}{MISE(f_n, h) + MISE(f_n, h')} \rightarrow 0 \quad \text{a.s.} \quad (5.8)$$

To prove (5.8) it is enough to show that

$$\sup_h \frac{|CV(h) - ISE(f_n, h) - 2[n^{-1} \sum_{i=1}^n \frac{f^0(X_i) w(X_i)}{H_n^*(X_i)} 1_{[\Delta_i=1]} - \int (f^0(x))^2 w(x) dx]|}{MISE(f_n, h)} \rightarrow 0 \quad \text{a.s.}$$

This may be rewritten as

$$\sup_h \frac{|2n^{-1}(n-1)^{-1} \sum_{i=1}^n \sum_{j \neq i} U_{ij}|}{MISE(f_n, h)} \rightarrow 0 \quad \text{a.s.,}$$

where

$$U_{ij} = h^{-1} K\left(\frac{X_i - X_j}{h}\right) \frac{w(X_i)}{H_n^*(X_i) H_n^*(X_j)} 1_{[\Delta_i=1, \Delta_j=1]} \\ - \int h^{-1} K\left(\frac{x - X_j}{h}\right) \frac{f^0(x) w(x)}{H_n^*(X_j)} dx 1_{[\Delta_j=1]}$$

$$\begin{aligned}
& - \frac{f^0(X_i)w(X_i)}{H_n^*(X_i)} 1_{[\Delta_i=1]} + \int [f^0(x)]^2 w(x) dx \\
& = U'_{ij} + Z_{ij},
\end{aligned}$$

defining

$$\begin{aligned}
U'_{ij} &= h^{-1} K\left(\frac{X_i - X_j}{h}\right) \frac{w(X_i)}{H_n^*(X_j)H_n^*(X_i)} 1_{[\Delta_i=1, \Delta_j=1]} \\
& - \int h^{-1} K\left(\frac{x - X_j}{h}\right) \frac{f^0(x)w(x)}{H_n^*(X_j)} dx 1_{[\Delta_j=1]} \\
& - \frac{f^0(X_i)w(X_i)}{H_n^*(X_i)} 1_{[\Delta_i=1]} + \int [f^0(x)]^2 w(x) dx
\end{aligned}$$

and

$$\begin{aligned}
Z_{ij} &= \left[h^{-1} K\left(\frac{X_i - X_j}{h}\right) \frac{w(X_i)}{H_n^*(X_j)} 1_{[\Delta_i=1, \Delta_j=1]} \right. \\
& \left. - f^0(X_i)w(X_i) 1_{[\Delta_i=1]} \right] \left[\frac{1}{H_n^*(X_i)} - \frac{1}{H_n^*(X_j)} \right].
\end{aligned}$$

Theorem 2.2 then follows from the two lemmas:

Lemma 3.

$$\sup_h \frac{\left| n^{-1}(n-1)^{-1} \sum_{i=1}^n \sum_{j \neq i} U'_{ij} \right|}{\text{MISE}(\hat{f}, h)} \rightarrow 0 \quad \text{a.s.}$$

Lemma 4.

$$\sup_h \frac{\left| n^{-1}(n-1)^{-1} \sum_{i=1}^n \sum_{j \neq i} Z_{ij} \right|}{\text{MISE}(\hat{f}, h)} \rightarrow 0 \quad \text{a.s.}$$

Proof of Lemma 3. This proof combines the ideas of Lemma 2 of Marron (1985) with those of the proof of Lemma 1 above. Recall that in the proof of Lemma 1, the notation E_{Δ_i} meant expected value taken over $\{X_i: \Delta_i=1\}$, conditioned on

the event $\{N=v\}$. The censored observations $\{X_i: \Delta_i=0\}$ were ignored in the definition of E_v since they did not appear in the quantities being analyzed. The censored X_i 's do appear in the following, so it will be understood that E_v denotes expected value as above, only also conditioned on $\{X_i: \Delta_i=0\}$ (or, equivalently, E_v denotes integration over $\{X_i: \Delta_i=1\}$, which are i.i.d random variables with density $f^0 H^*/p$).

For $v = 1, \dots, n$, $U'_{ij} = U''_{ij} + Z'_{jv}$, where

$$\begin{aligned} U''_{ij} &= h^{-1} K\left(\frac{X_i - X_j}{h}\right) \frac{w(X_i)}{H_n^*(X_j)H_n^*(X_i)} 1_{[\Delta_i=1, \Delta_j=1]} \\ &\quad - \frac{v}{np} \int h^{-1} K\left(\frac{x - X_j}{h}\right) \frac{w(x)}{H_n^*(x)} f^0(x) dx 1_{[\Delta_j=1]} \\ &\quad - \frac{f^0(X_i)w(X_i)}{H_n^*(X_i)} + \frac{v}{np} \int [f^0(x)]^2 w(x) dx \end{aligned}$$

and

$$\begin{aligned} Z'_{jv} &= \left(\frac{v}{np} - 1\right) \left[\int h^{-1} K\left(\frac{x - X_j}{h}\right) \frac{w(x)}{H_n^*(x)} f^0(x) dx 1_{[\Delta_j=1]} \right. \\ &\quad \left. - \int [f^0(x)]^2 w(x) dx \right]. \end{aligned}$$

Using the method of proof of Lemma 2 of Marron (1985), it can be shown that, for $k=1,2,\dots$, and n sufficiently large,

$$\sup_h E_v \left[\frac{n^{-1}(n-1)^{-1} \sum_{i=1}^n \sum_{j \neq i}^n U''_{ij}}{\text{MISE}(\hat{f}, h)} \right]^{2k} \leq C n^{-\gamma k},$$

regardless of the realization of $\{X_i: \Delta_i=0\}$. In a similar manner (i.e., approximate $H_n^*(x)$ by $H^*(x)$, including another $\frac{v}{np} - 1$ term, and using the cumulant-style argument of Marron (1985)), we can obtain

$$\sup_h E_v \left[\frac{n^{-1} \sum_{j=1}^n Z'_{jv}}{\text{MISE}(\hat{f}, h)} \right]^{2k} \leq C n^{-\gamma k}.$$

These two inequalities may now be used in a computation similar to that yielding (5.6) in the proof of Lemma 1 to finish the proof of Lemma 3.

Proof of Lemma 4. Write

$$\begin{aligned}
 & |n^{-1}(n-1)^{-1} \sum_i \sum_{j \neq i} z_{ij}| \\
 &= |n^{-1} \sum_i [f_{ni}(X_i) - f^0(X_i)] [\frac{1}{H_n^*(X_i)} - \frac{1}{H^*(X_i)}] 1_{[\Delta_i=1]} w(X_i)| \\
 &\leq \{n^{-1} \sum_i [f_{ni}(X_i) - f^0(X_i)]^2 1_{[\Delta_i=1]} w(X_i)\}^{1/2} \\
 &\quad \times \{n^{-1} \sum_i [\frac{1}{H_n^*(X_i)} - \frac{1}{H^*(X_i)}]^2 1_{[\Delta_i=1]} w(X_i)\}^{1/2}. \quad (5.9)
 \end{aligned}$$

The expression inside the first square-root on the right-hand side of (5.9) is the "leave-one-out" version of the average squared error and will be denoted by $ASE(f_{ni})$. Using the methods of Lemma 1 of Marron (1985) and Theorem 2 of Marron and Härdle (1986), it can be shown that for $k=1,2,\dots$ there is a constant C so that

$$E \left[\frac{ASE(f_{ni}) - MISE(f_n, h)}{MISE(f_n, h)} \right]^{2k} \leq C n^{-\gamma k}.$$

The proof of Lemma 4 is then completed by a computation like that leading to (5.6) in the proof of Lemma 1, which includes the uniform convergence result for the product-limit estimator H_n^* used in the proof of Lemma 2.

REFERENCES

- Blum, J. R. and Susarla, V. (1980), "Maximal deviation theory of density and failure estimates based on censored data," Multivariate Analysis-V, P. R. Krishnaiah, Ed., North-Holland, New York, 213-222.
- Breslow, N. and Crowley, J. (1974), "A large sample study of the life table and product-limit estimates under random censorship," Annals of Statistics, 2, 437-453.
- Bowman, A. (1984), "An alternative method of cross-validation for the smoothing of density estimates," Biometrika, 71, 353-360.
- Bretagnolle, J. and Huber, C. (1979). "Estimation des densites: Risque minimax," Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 47, 119-137.
- Burman, P. (1985), "A data dependent approach to density estimation," Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 69, 609-628.
- Csörgő, S. and Horváth, L. (1983), "The rate of strong uniform consistency for the product-limit estimator," Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 62, 411-426.
- Diehl, S. and Stute, W. (1985), "Kernel density estimation in the presence of censoring," Technical Report, University of Giessen.
- Efron, B. (1967), "The two-sample problem with censored data," Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 4, 831-853.
- Földes, A. and Rejtő, L. (1981), "A LIL type result for the product-limit estimator," Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 56, 75-86.
- Földes, A., Rejtő, L. and Winter, B. B. (1980), "Strong consistency properties of nonparametric estimators for randomly censored data, I: The product-limit estimator," Periodica Mathematica Hungarica, 11, 233-250.
- Földes, A., Rejtő, L. and Winter, B. B. (1981), "Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate," Periodica Mathematica Hungarica, 12, 15-29.
- Gill, R. (1983), "Large sample behavior of the product-limit estimator on the whole line," Annals of Statistics, 11, 49-58.
- Hall, P. (1983), "Large sample optimality of least squares cross-validation in density estimation," Annals of Statistics, 11, 1156-1174.
- Hall, P. (1985), "Asymptotic theory of minimum integrated square error for multivariate density estimation," Proceedings of the Sixth International Symposium on Multivariate Analysis at Pittsburgh, 25-29.

- Hall, P. and Marron, J. S. (1985a), "Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation," Center for Stochastic Processes Technical Report No. 94.
- Hall, P. and Marron, J. S. (1985b), "The amount of noise inherent in bandwidth selection for a kernel density estimator," Center for Stochastic Processes Technical Report No. 100.
- Kaplan, E. L. and Meier, P. (1958), "Nonparametric estimation from incomplete observations," Journal of the American Statistical Association, 53, 457-481.
- Marron, J. S. (1985), "A comparison of cross-validation techniques in density estimation," N. C. Institute of Statistics Mimeo Series #1568.
- Marron, J. S. and Härdle, W. (1986), "Random approximations to some measures of accuracy in nonparametric curve estimation," Journal of Multivariate Analysis, (to appear).
- McNichols, D. T. and Padgett, W. J. (1986), "Mean and variance of a kernel density estimator under the Koziol-Green model of random censorship," Sankhya, Series A, (to appear).
- Padgett, W. J. and McNichols, D. T. (1984), "Nonparametric density estimation from censored data," Communications in Statistics-Theory and Methods, 13, 1581-1611.
- Parzen, E. (1962), "On estimation of a probability density function and mode," Annals of Mathematical Statistics, 33, 1065-1076.
- Rosenblatt, M. (1956), "Remarks on some nonparametric estimates of a density function," Annals of Mathematical Statistics, 27, 832-837.
- Rosenblatt, M. (1971), "Curve estimates," Annals of Mathematical Statistics, 42, 1815-1842.
- Rosenblatt, M. (1976), "On the maximal deviation of k-dimensional density estimates," Annals of Probability, 4, 1009-1015.
- Rudemo, M. (1982), "Empirical choice of histograms and kernel density estimators," Scandinavian Journal of Statistics, 9, 65-78.
- Stone, C. J. (1984), "An asymptotically optimal window selection rule for kernel density estimates," Annals of Statistics, 12, 1285-1297.
- Stute, W. (1985), "Optimal bandwidth selection in pointwise density and hazard rate estimation when censoring is present," Technical Report, University of Giessen.
- Susarla, V., Tsai, W. Y. and Van Ryzin, J. (1984), "A Buckley-James-type estimator for the mean with censored data," Biometrika, 71, 624-625.

Watson, G. S. and Leadbetter, M. R. (1963), "On the estimation of the probability density, I," Annals of Mathematical Statistics, 34, 480-491.

Wellner, J. (1982), "Asymptotic optimality of the product-limit estimator," Annals of Statistics, 10, 595-602.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Statistics Technical Report No. 113	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Asymptotically Optimal Bandwidth Selection for Kernel Density Estimators from Randomly Right- Censored Samples		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Stat. Tech. Rpt. No. 113
7. AUTHOR(s) J. S. Marron and W. J. Padgett		8. CONTRACT OR GRANT NUMBER(s) MIPR ARO 139-85
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of South Carolina Columbia, SC 29208		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304A5 N/A
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE March, 1986
14. MONITORING AGENCY NAME & ADDRESS (If different from Controlling Office)		13. NUMBER OF PAGES 24
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) NA		
18. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Nonparametric density estimation; Optimal bandwidth; Random censorship; Smoothing parameter; Cross-validation.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper makes two important contributions to the theory of bandwidth selection for kernel density estimators under right censorship. First, an asymptotic representation of the integrated squared error into easily understood variance and squared bias components is given. Second, it is shown that if the bandwidth is chosen by the data-based method of least squares cross-validation, then it is asymptotically optimal in a compelling sense. A by-product of the first part is an interesting comparison of the two most popular kernel estimators.		

END

DTIC

8-86